

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 October 2003 (09.10.2003)

PCT

(10) International Publication Number
WO 03/084172 A1

(51) International Patent Classification⁷: **H04L 29/06**,
H04N 7/24

(21) International Application Number: PCT/GB03/01358

(22) International Filing Date: 27 March 2003 (27.03.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02252224.7 27 March 2002 (27.03.2002) EP

(71) Applicant (for all designated States except US): **BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY** [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **NILSSON, Michael**, Erling [GB/GB]; 5 Laburnum Gardens, Rushmere St Andrews, Ipswich, Suffolk IP5 1BY (GB). **JEBB, Timothy, Ralph** [GB/GB]; 1 Fairbairn Avenue, Kesgrave, Ipswich, Suffolk IP5 2YS (GB).

(74) Agent: **LLOYD, Barry, George, William**; BT Group Legal, Intellectual Property Department, Holborn Centre, 8th Floor, 120 Holborn, London EC1N 2TE (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

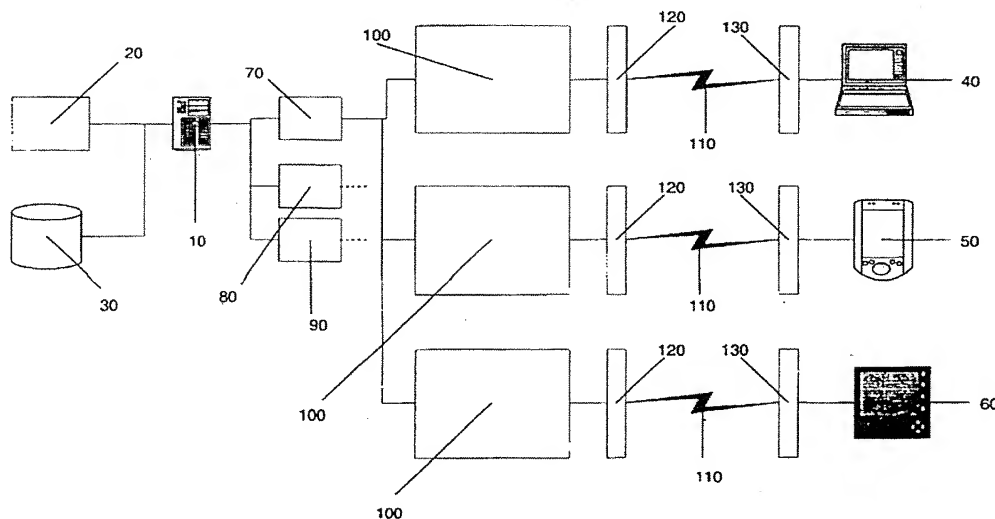
(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

[Continued on next page]

(54) Title: DATA STREAMING SYSTEM AND METHOD



(57) Abstract: A data streaming system and method are described. A server (10) is arranged to stream one of a plurality of encoded data streams to a client (40, 50, 60). Each of the plurality of data streams is an independent representation of a common data source encoded at a different resolution to the other of the plurality of data streams. The server (10) comprises a transmitter (100) and a first buffer (120). The transmitter is arranged to transmit data packets of the encoded data stream to the client (40, 50, 60) via the first buffer (120). The transmitter (100) is arranged to monitor the content of the first buffer (120) and switch to transmit another of the plurality of data streams in the event that predetermined criteria are detected from the first buffer (120).

WO 03/084172 A1



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Data Streaming System and Method

The present invention relates to a system and method suitable for streaming audio and video content over IP (Internet Protocol) networks. In particular, the present invention is suitable for use where the available bit rate is inherently variable due to physical network characteristics and/or contention with other traffic. For example, the present invention is suitable for multimedia streaming to mobile handheld terminals, such as PDAs (Personal Digital Assistants) via GPRS (General Packet Radio Service) or 3G networks.

New data network access technologies such as cable and ADSL (Asymmetric Digital Subscriber Line) modems, together with advances in compression and the availability of free client software are driving the growth of video streaming over the Internet. The use of this technology is growing exponentially, possibly doubling in size every six months, with an estimated half a billion streams being served in 2000. However, user perception of Internet streaming is still coloured by experiences of congestion and large start-up delays.

Current IP networks are not well suited to the streaming of video content as they exhibit packet loss, delay and jitter (delay variation), as well as variable achievable throughput, all of which can detract from the end-user's enjoyment of the multimedia content.

Real-time video applications require all packets to arrive in a timely manner. If packets are lost, then the synchronisation between encoder and decoder is broken, and errors propagate through the rendered video for some time. If packets are excessively delayed, they become useless to the decoder, which must operate in real-time, and are treated as lost. Packet loss, and its visual effect on the rendered video, is particularly significant in predictive video coding systems, such as H.263. The effect of packet loss can be reduced, but not eliminated, by introducing error protection into the video stream. It has been found that such resilience techniques can only minimise, rather than eliminate, the effect of packet loss.

In the case of a sustained packet loss, indicating a long-term drop in throughput, the streaming system needs to be able to reduce its long term requirements. This commonly means that the bit-rate of the streamed media must be reduced.

5

Standard compression technologies, such as H.263 and MPEG-4, can be managed to provide a multimedia source that is capable of changing its encoding rate dynamically. A video source having such properties is described herein as an elastic source, i.e. one that is capable of adapting to long-term variations in network throughput. This is commonly achieved by providing a continuously adaptive video bit-rate. This is possible because unlike audio codecs, video compression standards do not specify an absolute operating bit-rate.

10

Video streaming systems may be designed to provide an encoded stream with varying bit rate, where the bit rate adapts, in response to client feedback, instantly to the available network bandwidth. Such a system could be made to be network-friendly, by controlling the transmission rate such that it reduces rapidly in the case of packet loss, and increases slowly at other times.

15

However, this solution is not practical for two reasons. Firstly, real-time video encoding usually requires a large amount of processing power, thus preventing such a solution from scaling to support many users. Secondly, the end-user perception of the overall quality will be adversely affected by rapid variations in instantaneous quality.

20

For uni-directional streaming applications, the delay between the sender and receiver is only perceptible at start-up. Therefore, common techniques trade delay for packet loss and jitter. Provided the average throughput requirements of the video stream match the average available bandwidth the receiver buffer size can be dimensioned to contain the expected variation in delay.

25

30

Market-leading streaming systems are believed to use significant client-side buffering to reduce the effects of jitter that may be encountered in the Internet. While this helps,

it also introduces large start-up delays, typically between 5 and 30 seconds, as the buffer fills. These systems also include technologies that allow the client to adapt to variations in available bandwidth. Although the details of these techniques are not publicly available, it is suspected that they generally use multi-data rate encoding within single files (SNR scalability), and intelligent transmission techniques such as server-side reduction of the video picture rate to maintain audio quality. Such large amounts of buffering could conceivably allow a significant proportion of packets to be resent, although these re-transmissions themselves are subject to the same network characteristics. The decision to resend lost data is conditional on this and several other factors. Such techniques are generally only applicable to unicast transmissions. Multicast transmission systems are typically better served by forward error correction or receiver-based scalability such as RLM and RLC. S. McCanne, 'Receiver driven layered multicast', Proceedings of SIGCOMM 96, Stanford. CA. August 1996. L. Vicisano, L. Rizzo and J. Crowcroft, 'TCP-like congestion control for layered multicast data transfer', Infocom '98.

The use of a buffer as described above allows a system to overcome packet loss and jitter. However, it does not overcome the problem of there being insufficient bit rate available from the network. If the long term average bit rate requirements of the video material exceeds the average bit rate available from the network, the client buffer will eventually be drained and the video renderer will stop until the buffer is refilled. The degree of mismatch between available network bit rate and the rate at which the content was encoded determines the frequency of pausing to refill the buffer.

As described above, most video compression algorithms, including H.263 and MPEG-4, can be implemented to provide a continuously adaptive bit rate. However, once video and audio have been compressed, they become inelastic, and need to be transmitted at the encoded bit-rate.

Whilst network jitter and short term variations in network throughput can be absorbed by operating a buffer at the receiver, elasticity is achieved only when long-term variations in the network throughput can also be absorbed.

Layered encoding is a well-known technique for creating elastic video sources. Layered video compression uses a hierarchical coding scheme, in which quality at the receiver is enhanced by the reception and decoding of higher layers, which are sequentially added to the base representation. At any time, each client may receive any number of these video layers, depending on their current network connectivity to the source. In its simplest implementation, this provides a coarse-grain adaptation to network conditions, which is advantageous in multicast scenarios. Layered video compression has also been combined with buffering at the client, to add fine-grain adaptation to network conditions. However, it has been shown that layered encoding techniques are inefficient, and will typically require significantly more processing at the client which causes particular problems when dealing with mobile devices, which are likely to have reduced processing capability.

Transcoding is another well-known technique for creating elastic video sources. It has been shown that video transcoding can be designed to have much lower computational complexity than video encoding. However, the computational complexity is not negligible, and so would not lead to a scalable architecture for video streaming.

According to one aspect of the present invention, there is provided a data streaming system comprising a server arranged to stream one of a plurality of encoded data streams to a client, each of the plurality of data streams being an independent representation of a common data source encoded at a different resolution to the other of the plurality of data streams, the server comprising a transmitter and a first buffer, the transmitter being arranged to transmit data packets of the encoded data stream to the client via the first buffer, wherein the transmitter is arranged to monitor the content of the first buffer and switch to transmit another of the plurality of data streams in the event that predetermined criteria are detected from the first buffer.

Some of the key attributes of the overall system are:

- varying the transmission rate in a network-friendly manner;
- decoupling of the transmission rate from the media encoding rate;

- building up a buffer of data at the client without incurring a start-up delay;
- smoothing short term variations in network throughput by use of client buffering;
- adjusting long-term average bandwidth requirements to match the available resources in the network by switching between multimedia streams encoded at different bit rates; and,
- providing resilience to packet loss by selectively retransmitting lost packets, without affecting the quality perceived by the user, by use of client buffering.

10 The present invention permits scaling the transmission bit rate of the compressed video in dependence on changing network conditions.

In the present invention, a produced audio-visual stream does not have to be transmitted at a single fixed bit rate, thus allowing transmission at whatever rate the network
15 instantaneously supports. Resilience to transmission losses is provided by building a buffer of data at the receiver, to allow time for lost data to be retransmitted before it is needed for decoding and presentation.

At any one time, only one video stream and one audio stream from a hierarchy of such
20 streams are transmitted to a client. This is implemented in the form of a combination of so called "simulcast switching" for coarse-grain adaptability, and transmission rate variation for fine-grain adaptation.

The system has been shown to perform well over a GPRS network, making good use of
25 the available network bandwidth, to provide satisfactory multimedia quality.

The system has been designed to overcome the characteristics of IP networks, and in particular mobile IP networks, to provide users with multimedia of consistent quality with minimal start-up delay.

30

The transmitter may be arranged to determine the amount of data buffered at the client from the content of the first buffer, wherein the predetermined criteria include a

6

predetermined level of data determined to be buffered at the client. A data packet may be removed from the first buffer upon acknowledgement by the client of receipt of the packet. The transmitter may be arranged to determine the amount of data buffered at the client in dependence on the latest data packet removed from the first buffer and on
5 an estimation of number of packets decoded by the client.

The first buffer may include a mirror buffer storing data on packets in the first buffer, the transmitter being arranged to monitor the content of the first buffer using the data in the mirror buffer.

10 Data packets may be transmitted to the client using an extended TPKT protocol, the data packets including a header containing a decoding timestamp and a data stream identifier.

15 The system may further comprise a plurality of transmitters, each communicating with a respective client via a respective first buffer to transmit one of the plurality of data streams determined in dependence on respective predetermined criteria.

The data stream may be encoded video data.

20 The transmitter may be arranged to multiplex audio packets and video packets within the transmission of data packets. Neighbouring audio and video packets may represent audio and video information that is intended for representation at substantially the same time.

25 The data stream may be encoded audio data.

The resolution may be an encoding bit rate of the data.

30 The server may include an encoder arranged to accept a data feed and encode the data feed into the plurality of encoded data streams.

The system may further comprise a plurality of buffers, wherein the encoder is arranged to output each encoded data stream into a respective one of the plurality of buffers, the transmitter being arranged to obtain data packets for a respective data stream from its respective one of the plurality of buffers.

5

The server may include a file source storing the plurality of encoded data streams.

According to another aspect of the present invention, there is provided a data streaming system comprising a client and a server, the server being arranged to stream one of a plurality of encoded data streams to the client, each of the plurality of data streams being an independent representation of a common data source encoded at a different resolution to the other of the plurality of data streams, the server comprising a transmitter and a first buffer and the client including a receiving buffer, wherein the transmitter is arranged to transmit data packets of the encoded data stream to the client via the first buffer, wherein the client is arranged to store received data packets in the receiving buffer and to acknowledge receipt to the server, wherein the transmitter is arranged to delete packets from the first buffer when an acknowledgement receipt is received, the server being arranged to switch to another of the plurality of data streams in the event that predetermined criteria are satisfied, the predetermined criteria comprising analysis on content of the first buffer.

10
15
20

The packets may include packet sequence data, the client being arranged to request retransmission of non-received packets based on the sequence data, the server being arranged to retransmit a packet from the first buffer upon receipt of a retransmission request.

25

According to a further aspect of the present invention, there is provided a method of streaming one of a plurality of encoded data streams to a client, each of the plurality of data streams being an independent representation of a common data source encoded at a different resolution to the other of the plurality of data streams, the method comprising the steps of:
transmitting data packets of the encoded data stream to the client via a first buffer;

30

8

monitoring the content of the first buffer; and,
switching to transmit another of the plurality of data streams in the event that predetermined criteria are detected from the first buffer.

- 5 The plurality of data streams may each be encoded at a different bit rate, the method further comprising the step of initially transmitting data packets of the lowest bit rate data stream.

The predetermined criteria may include an amount of data determined to be buffered at
10 client.

The predetermined criteria may include one or more network throughput thresholds.

Network throughput may be calculated by the steps of:

- 15 counting the number of bytes passed to the first buffer;
subtracting the counted number of bytes from the size of the first buffer; and,
dividing the result by the time since the start of transmission.

The method may further comprise the step of measuring network throughput over more
20 than one interval to determine throughput variation.

The predetermined criteria may include determination of network throughput sufficient to sustain the other of the plurality of the data streams.

- 25 The method may further comprise the step transmitting data at a maximum rate irrespective of an amount of data buffered at the client, wherein the predetermined criteria include network throughput determined at the maximum rate.

The data streams may be encoded as a series of pictures predictively encoded in
30 dependence on the previous pictures in the data stream, the data streams including quantised source access pictures interspersed at predetermined periods in the picture

series, wherein the method of encoding the quantised source access pictures including the steps of:

encoding picture as a predicted picture; and,

- if no information about an area of a picture is indicated in the encoded predicted picture, setting the quantiser index to a fine quantisation value when encoding as a quantised source access picture.

Examples of the present invention will now be described in detail, with reference to the accompanying Figures, in which:

- Figure 1 is a schematic diagram of an audio-visual data streaming system in accordance with an embodiment of the present invention;

Figure 2 is a schematic diagram of a video encoding hierarchy used in the system of Figure 1.

- Figure 3 is a schematic diagram of a video encoding architecture that allows mismatch free switching between video streams to be achieved.

Figure 4 is a schematic diagram of a client-server architecture suitable for use in the system of Figure 1;

- Figures 5a and 5b are, respectively, diagrams illustrating standard TKPT transport packet structure and a variation of that structure implemented for the present invention; and,

Figures 6a-6c are schematic diagrams illustrating aspects of a data structure comprising an audio-visual data stream suitable for storing data for use in the present invention.

- Figure 1 is a schematic diagram of an audio-visual data streaming system in accordance with an embodiment of the present invention.

- The server 10 receives encoded multimedia content either directly from an encoder 20 or from a file 30, and serves this content to one or more clients 40-60. The server 10 scales to support many clients 40-60 accessing many pieces of content independently as it performs little processing, just selecting packets for onward transmission. No encoding or transcoding of media is performed in the server 10.

10

In principle, the server 10 operates in the same way for both live streams, provided from the encoder 20, and for pre-encoded streams from the file 30. In this particular embodiment, streaming of live media is described. Differences in streaming media from pre-encoded files are discussed in later embodiments.

5

The server 10 includes a number of circular buffers 70-90. For each client 40-60 there is one instance of a packet transmitter 100. The packet transmitter 100 determines when and from which buffer 70-90 the next packet is read, reads the chosen packet and sends it to the respective client over a network connection 110.

10

A semi-reliable network connection 110 is required from the server 10 to each respective client 40-60 to ensure that almost all packets sent are received, therefore minimising disturbances to user-perceived quality. Buffers (120, 130) are therefore used at the respective ends of the network connection 110 to allow retransmissions of lost packets. The network connection 110 is also desired to be network friendly, that is, to allow the bit rate used to be increased when congestion is not experienced, and to be drastically reduced when congestion occurs.

15

Whilst the system components are illustrated and described as a combination of integrated and separate components, it will be appreciated that different configurations could be used. For example, an external encoder 20 and/or file store 30 could be used. Equally, the buffers 130 are likely to be integral to the client devices 40-60.

20

Figure 2 is a schematic diagram of a video encoding hierarchy used in the system of Figure 1. The encoder 20 encodes live or stored multimedia content into an elastic encoded representation. Audio is encoded at low bit rate into a single encoded bit stream, and hence is in itself inelastic. However, as audio typically requires a smaller bit rate than video, provided the video is encoded in an elastic fashion, then the combined encoding of audio and video can be considered to be elastic.

25

30

Audio is encoded using the AMR (Adaptive Multi-Rate) encoder at 4.8 kbit/s. Video is encoded into an elastic representation. In a manner similar to layering, the encoder 20

11

creates a hierarchy of independent video streams. Instead of building this hierarchy by making each stream dependent on all streams lower in the hierarchy, each stream is encoded independently. Such a hierarchy is well-known, being referred to as 'simulcast'.

5

Although audio data has been described as being encoded using a low bit rate AMR scheme, other AMR encoding rates, and other encoding standards such as MP3, could also be supported. Encoded audio at various rates could be organised in a hierarchy of independent streams in a similar manner to that described below for video, but with the simplification of switching between encoded representations from the fact that each audio frame is typically coded independently.

10

The video hierarchy, created using an extension to the ITU-T standard H.263, includes an intra stream 200, to allow random access to video streams, and one or more play streams 210a, 210b, for ordinary viewing of the content. Each play stream 210a, 210b is encoded at a different bit rate, thus allowing a given client 40-60 to receive at a rate appropriate for its current network connection 110 to the server 10. The hierarchy also contains switching streams 220, 230, 240 which allow switching from the intra stream 200 to the lowest rate play stream 210a, and between play streams.

15

20

Since the encoding algorithms employ motion-compensated prediction, switching between bitstreams at arbitrary points in a play stream, although possible, would lead to visual artifacts due to the mismatch between the reconstructed frames at the same time instant of different bit streams. The visual artifacts will further propagate in time.

25

In current video encoding standards, perfect (mismatch-free) switching between bit streams is possible only at the positions where the future frames/regions does not use any information previous to the current switching location, i.e., at access pictures. Furthermore, by placing access pictures at fixed (e.g. 1 sec) intervals, VCR functionalities, such as random access or "Fast Forward" and "Fast Backward" (increased playback rate) for streaming video content, are achieved. A user can skip a portion of video and restart playing at any access picture location. Similarly, increased

30

playback rate, i.e., fast-forwarding, can be achieved by transmitting only access pictures.

It is, however, well known that access Pictures require more bits than the motion-compensated predicted frames. Thus, the intra stream 200 and switching streams 220, 230, 240 are used. The main property of switching streams is that identical pictures can be obtained even when different reference frames are used.

The main purpose of the hierarchy is to allow the server 10 to transmit a play stream 210a or 210b to a client 40-60 to achieve an optimal balance between building up a buffer of received data at the client 40-60 to provide resilience to packet loss and sudden drops in network throughput, and providing the best play stream 210a or 210b to the client 40-60 depending on the highest bit rate that its network connection 110 instantaneously supports.

The intra stream 200 is a series of intra coded pictures (201, 202) that are used to provide random access and recovery from severe error conditions. The play streams 210a, 210b include predictively coded pictures (211a, 212a, 213a, 214a, 215a; 211b, 212b, 213b, 214b, 215b) which may be bi-directionally predicted, and may be predicted from multiple reference pictures. The play streams 210a, 210b also include periodic access Pictures 216a, 217a; 216b, 217b. The switching streams 220, 230, 240 consist of a series of linking Pictures (221, 222; 231, 232; 241, 242).

The circular buffers 70-92 are designated for each stream type, one for each intra (70), play (80, 85) and switching (90, 91, 92) stream for each piece of content

When a client 40 first connects to the server 10, the server 10 locates an appropriate intra picture (for example, intra picture 201) from the circular buffer 70 storing the intra stream, and sends this to the client 40. The server 10 then selects the linking picture (221) to switch from the intra stream 220 to the play stream 210a with the lowest encoding bit rate, and then continue to serve from that play stream (213a onwards).

The transmission of packets to the client 40 is an independent process, with the rate of transmission depending on the state of the network and the transmission protocol used. However, the intention is that initially the transmission rate is greater than the encoding bit rate of the play stream 210a with the lowest encoding bit rate. This will allow the client 40 to start decoding and presenting media to the user immediately at the point that data is received and decoded, while also allowing the client 40 to build up excess compressed media data in its decoding buffer.

At the point where an access picture (such as access picture 217a in the above example), the client 40 and/or server 10 may determine that a different play stream is more suitable (for example due to increased or decreased network capacity). In the above example, switching from the low rate play stream 210a to the higher rate play stream 210b is accomplished by the server 10 transmitting the link picture 232 instead of access picture 217a. The link picture 232 links to play stream picture 215b of the higher rate play stream 210b allowing the client 40 to receive that play stream. Switching to a play stream of decreased bit rate is accomplished in a similar manner.

Three methods of encoding linking pictures have been investigated. Each method provides different compromises between the accumulation of drift from switching, the cost in terms of bit rate of the actual switching, and the impact on the quality of the individual play streams caused by encoding regular pictures of a type that allow drift-free low bit rate switching.

1. Predictively coded linking pictures
- In the first method, linking pictures are generated as Predicted pictures. They are coded in a manner such that when reconstructed they are similar, in the sense of having for example a small mean square difference, to the reconstruction of the simultaneous access picture in the destination play stream. Access pictures can be coded as Predicted pictures. The number of bits used to encode the linking pictures determine how well matched the reconstructed linking picture is to the reconstructed access picture, and hence determines the amount of drift that would

occur as a result of switching. However, drift will accumulate on each occurrence of switching.

2. Intra coded linking pictures

5 In the second method, linking pictures are generated as intra pictures. They are coded in a manner such that when reconstructed they are similar, in the sense of having for example a small mean square difference, to the reconstruction of the simultaneous access picture in the destination play stream. Access pictures can be coded as Predicted pictures. The number of bits used to encode the linking
10 pictures determines how well matched the reconstructed linking picture is to the reconstructed access picture, and hence the amount of drift that would occur as a result of switching. However, for a given amount of mismatch, an intra coded linking picture would usually require many more bits than a predictively coded linking picture. The use of intra coding for linking pictures prevents the
15 accumulation of drift.

3. Quantised-Source coded linking pictures

In the third method, linking pictures are coded with a technique based on the concept described in "VCEG-L27, A proposal for SP-frames, submitted by Marta
20 Karczewicz and Ragip Kurceren at the ITU-Telecommunications Standardization Sector Video Coding Experts Group's Twelfth Meeting: Eibsee, Germany, 9-12 January, 2001, available at <ftp://standard.pictel.com/video-site/>" referred to herein as Quantised-Source pictures.

25 The encoding architecture for Quantised-Source pictures is shown in Figure 3. The source picture and the motion compensated prediction are independently quantised in steps 300 and 310 respectively, with the same quantiser index, and transformed, before being subtracted in step 320 and variable length encoded in step 330. The reconstructed picture is formed by adding, in step 340, the output
30 of subtractor 320 and the output of quantisation and transformation 310, and inverse transforming and inverse quantising the result in step 350. The reconstructed picture is stored in Picture Store 360. The result is that the

reconstructed picture is simply the quantised source picture, and is independent of the motion compensated prediction. Hence a given source picture can be reconstructed identically when predicted from different reference pictures, and hence drift free switching is enabled. The motion compensated prediction is not irrelevant, as it reduces the entropy of the signal to be variable length encoded and hence reduces the number of bits produced by encoding a picture.

Access pictures are also coded as Quantised-Source pictures, with an identical selection of coding modes, intra or inter, and quantiser choice, as the linking picture. This ensures that the linking picture reconstructs identically to the simultaneous access picture in the destination play stream.

The number of bits required to encode the linking pictures is determined by the encoding of the corresponding access picture. The number of bits used to encode the access picture depends on how the quantisation is performed, but in general is more than the number of bits used to encode Predicted pictures and less than the number of bits used to encode Intra pictures. This is because encoding is more efficient than intra encoding due to the use of prediction, but not as efficient as normal prediction due to the quantisation of the prediction error. Hence the use of Quantised-Source pictures allows drift free switching but at the expense of less efficient encoding of the play stream.

Quantised-Source pictures are encoded with the same H.263 syntax as predicted pictures, with the exception that they are distinguished from predicted pictures by setting the first three bits of MPPTYPE to the reserved value of "110".

The periodic encoding of Quantised-Source pictures can cause a beating effect in stationary areas of pictures. This is explained as follows. In normal predictive coding, stationary areas of the picture which have already been encoded as a reasonable representation of the source picture are not modified. In the encoding of such areas in Quantised-Source pictures, the prediction must be quantised, and if done with the quantiser index used for non-stationary areas of the picture,

16

makes the region change, possibly making it worse, but in any case, changing it. This changing is the beating effect.

5 This is overcome by noting that when the prediction for an area of the picture provides a good enough representation of the source, there is no need to transmit information, and hence change the area. So when an access picture is encoded as a Quantised-Source picture, a test is performed to determine whether information about the area would have been transmitted if the picture had been encoded as a Predicted picture rather than a Quantised-Source picture. If no information would
10 have been transmitted, the quantiser index used by the quantisation of steps 300 and 310 and inverse quantisation of step 350 is set to a small value, the output of subtractor 320, commonly known as the prediction error, is set to zero, thus this area of the newly reconstructed picture is equal to the corresponding area of the previous reconstructed picture quantised with a fine quantiser. In H.263 and
15 other standards, the range of quantiser index is from 1 (fine) to 31 (coarse). By referring to a small index, a value typically of 8 or less is meant. This minimises unnecessary changes to the reconstructed picture while minimising the amount of information that must be transmitted. There will however be a cost in bit rate in the corresponding linking picture, where the prediction error is unlikely to be
20 zero, but the same fine quantiser must be used.

Figure 4 is a schematic diagram of a client-server architecture suitable for use in the system of Figure 1.

25 The client 40 includes a network buffer 130, a decoding buffer 41 and a decoder 42. The server 10 includes circular buffers 70, 80, 90 as discussed above, and a packet transmitter 100 and network buffer 120 for each client.

The client 40 keeps the server 10 informed of the amount of information in its decoding
30 buffer 41 and the rate at which it is receiving data. The server 10 uses this information to determine when to switch between play streams. For example, when the client 40 has accumulated more than a threshold of data, say 15 seconds of data in its decoding

buffer 41 and the client 40 is receiving at a rate greater than or equal to the encoding rate of the next higher play stream in the hierarchy, the server 10 can switch the client's packet transmitter 100 to the next higher play stream at the next linking picture.

- 5 Similarly, when the amount of data accumulated by the client 40 in its decoding buffer 41 falls to less than a threshold, the server 10 can switch the client's packet transmitter 100 to the next lower play stream at the next linking picture.

- The overall effect is that the transmission rate varies in a network-friendly fashion
10 according to the state of congestion in the network, but due to the accumulation of data in the client's decoding buffer 41, the user perceives no change in quality as a result of short term changes in transmission rate. Longer term changes in transmission rate are handled by switching to a stream with a different encoding rate, to allow increased quality when the network allows it, and to reduce quality, without stalling presentation
15 or presenting corrupted media to the user, when the network throughput drops.

- The decoding buffer 41 at the client is used to reduce the impact of network performance variations on the quality of media presented to the user. The network characteristics that the buffer is designed to handle fall into three categories: packet
20 jitter, packet loss and variable throughput. In practice these three network characteristics are not independent, all being associated with network congestion, and in the case of mobile networks, with degradation at the physical layer.

- By de-coupling the transmission rate from the media encoding rate, the client's
25 decoding buffer 41 can be filled when network conditions are favourable, to provide resilience for times when network conditions are not so good.

- The accumulation of tens of seconds of data in the decoding buffer 41, allows packet jitter (delay variations) of the same magnitude to be masked from the user. In practice
30 this masks all packet jitter, as larger amounts of jitter are better classified as temporary connection drop-outs, which are handled by the error recovery process described below.

By accumulating data in the decoding buffer 41, time is available for the retransmission of lost packets before they are needed for decoding. Again, by dimensioning the decoder buffer 41 to contain more data than some multiple of the round trip delay, there is time for a small number of retransmission attempts to recover from packet loss. This allows recovery from most instances of packet loss without affecting decoded media quality, and makes the connection semi-reliable.

Finally, again by accumulating data in the decoding buffer 41, the client 40 can sustain consistent media quality for some time when the receiving bit rate is less than the encoding bit rate, and for some time when the receiving rate has dropped to zero.

As the data is streamed to the client 40 at a rate independent of the encoding rate, and buffered in the decoding buffer 41, it is necessary for decoding of data to be correctly timed, rather than simply to decode and present as fast as possible. Timestamps are used for this purpose, as well as for the synchronisation of audio and video.

Due to network variations, the amount of data in the client's decoding buffer 41, measured in bytes, may vary with time. In addition, the amount of data in the decoding buffer 41, measured in terms of the length of media presentation time it represents, would also vary with time. This has implications for streaming of live content: it is not possible to build up data in the decoding buffer 41 if the first data sent to the client 40 is sent with minimal delay from the time it was captured and encoded. Hence, the first data that is sent to the client 40 must be old data, that is, data representing events that took place some time before the client 40 connected to the server 10. Then as the decoding buffer 41 fills, the most recent data in it becomes more and more recent, while the media presented to the user remains at a constant delay from the actual time of occurrence.

The server buffers encoded data in its circular buffers 70, 80, 90, for a constant period of time after encoding so that when a client 40 connects to the server 10, 'old' data is available for streaming to the client 40. As the client's decoding buffer 41 fills, the

19

reading points from the circular buffers 70, 80, 90 get nearer to the newest data in these buffers.

The optimal sizing of the circular buffers 70, 80, 90, and the client decoding buffer 41,
5 is preferably such that each can contain the same amount of data, measured in terms of the media presentation time it represents.

The network buffers 120, 130 respectively in the server 10 and client 40 are used by a transport protocol implementing the semi-reliable data connection. Typically, data is
10 retained in the server's network buffer 120 until it, and all earlier data, have been acknowledged to have been received at the client 40. Similarly, data would be removed from the client's network buffer 130 when it, and all earlier data have been successfully received and passed to the decoding buffer 41. Consequently, the server 10, by knowing the data that remains in its own network buffer 120, knows what data has been
15 successfully received by the client 40, within bounds given by the uni-directional transmission delay.

This implies that no feedback from client 40 to server 10, beyond that needed by the transport protocol itself, is needed for the server 10 to know how much data has been
20 received by the client 40, so that it can make decisions about switching between play streams.

The presence of an accumulation of data in the client's decoding buffer 41 provides resilience to a number of network impairments, such as jitter, packet loss and variable
25 throughput. Clearly, it is not possible to recover from all network impairments unless the decoding buffer 41 is dimensioned to contain the whole media content and presentation is delayed until all data is received. As this case is not streaming, but downloading, a strategy to recover from serious network impairments is needed.

30 At times when the network throughput drops to a level below the encoding rate of the lowest rate play stream for a considerable length of time, the amount of data in the decoding buffer 41 will reduce and will eventually become zero. At this time,

20

presentation to the user will stop. However, circular buffer filling will continue at the server 10. Consequently, when the network recovers to a state in which transmission of the lowest rate play stream is again possible, the next data required by the client 40 will most likely not be in the server's circular buffer 70, 80, 90, as it would have been
5 overwritten by more recent data.

To recover from this situation, the server 10 must restart streaming as if a new connection had been made from the client: it must find a point in the intra stream, and start streaming from it, and then switch through the linking stream into the lowest rate
10 play stream. The effect on the user will be the loss of media from the time that the decoding buffer 41 became empty to the time when the server starts to send the intra stream.

The server 10 will be aware of the client's decoding buffer 41 becoming empty as it is
15 aware of when the client started to decode and of how much data has been successfully received. It will therefore be able to restart at an intra stream picture without the need for a specific message from the client. However, to provide resilience to the system, for example to recover from the effect of different clock speeds in the server and the client, a control message is sent from the client 40 to the server 10 in this situation.

20 In principle, streaming from file is identical to live streaming. In practice, it is somewhat simpler. There is no need for Circular Buffers 70, 80, 90 as data can be read from file as and when needed. The server 10 however uses the same techniques to fill up the decoding buffer 41 at the client 40 and to switch between play streams. In the
25 case of the decoding buffer 41 becoming empty, there is no need to restart at a later point in the content with an intra stream picture, as presentation can resume when the network throughput again becomes sufficient: the user simply perceives a period in which no media is presented.

30 Trick modes, such as fast forward, fast reverse and random access, become possible by use of the intra stream.

By writing 'old' data in the circular buffers 70, 80, 90 to file just before being overwritten, the problem described above of the decoding buffer 41 becoming empty, and the user missing content until recovery with an intra stream picture occurs, can be avoided, as data for streaming to the client will always be available: it will have to be
5 read from file rather than from the circular buffers 70, 80, 90.

Such functionality would also allow a client to pause the presented media for an indefinite period of time, and continue streaming afterwards. It would also allow the user to fast forward after such a pause to catch up with the live stream.

10

An implementation of the transport protocol tested in the above mentioned client-server architecture is based on the ISO TCP transport protocol TPKT, which is described in detail in RFC-2126 by Y. Pouffary, "ISO Transport Service on top of TCP (ITOT)".

15 The standard TPKT protocol defines a header illustrated in Figure 5a, followed by a payload. The packet length indicates the combined length of header and payload in octets.

In the implementation used for the present invention, TPKT is extended to have a
20 header, an example of which is illustrated in Figure 5b, followed by a payload. The packet length indicates the combined length of header, timestamp if present, and payload in octets. T is a bit that indicates whether the timestamp is present, and M is a bit that indicates whether the payload contains audio or video information.

25 As stated above, timestamps are required for the correct timing of decoding of data. Information embedded in packet headers include the length of the packet, a timestamp for the data in the packet, and a stream identifier.

The stream identifier is provided to allow audio and video to be multiplexed into a
30 single TCP connection. This is to ensure synchronisation of audio and video transmission. If separate TCP connections are used, it is possible that they will respond slightly differently to network characteristics and will achieve different throughputs,

which would result eventually in vastly different amounts of data in the client's decoding buffers, measured in terms of presentation time. Although these differences could be managed, the issue is totally avoided by using a single TCP connection and multiplexing audio and video with the same presentation time in neighbouring packets.

- 5 In fact, adding audio to a video only system simply requires the sending of audio packets at the same time as the associated video: no further control is necessary.

The server 10 attempts to send packets as quickly as possible. Initially, a number of packets are sent back-to-back regardless of the network capacity, as they are simply
10 building up in the server's network buffer 120. When the network buffer 120 becomes full, the rate at which packets can be sent to the network buffer 120 matches the rate of transmission over the network, with the transmission process being limited by blocking calls to the socket send function.

- 15 The transmission rate is also limited when the amount of data buffered at the client reaches a threshold, for example 30 seconds. When the client's decoding buffer 41 has this much data, the server 10 restricts the transmission rate to maintain this level of fullness.

- 20 Network throughput is estimated by counting bytes that have been sent to the network buffer 120, subtracting from this the size of the network buffer, and dividing by the time since the start of transmission. Shorter term estimates of network throughput are calculated using two counts of bytes transmitted and two measures of the time taken to send them, calculating the throughput from one pair, and switching between then
25 periodically, resetting the pair no longer being used to zero. For example, if resetting occurs every 200 seconds, the network throughput is estimated over a period that varies from 200 seconds immediately after resetting to 40 seconds just before resetting again.

- This technique works satisfactorily provided the server 10 is attempting to stream as
30 quickly as possible. But as mentioned above, when the amount of data in the decoding buffer 41 exceeds a threshold, the server 10 restricts its transmission rate to maintain a constant buffer fill. In this case, the network throughput would be estimated as the

encoding bit rate of the current play stream. When in this state, the network may be capable of transmitting a higher rate play stream than the one currently being streamed, but the server 10 does not switch because it can not make a true estimate of the network throughput because of its own rate limiting. To escape from this state, the server will

5 periodically ignore the client decoding buffer fullness threshold, and stream at full rate for a given period of time or given amount of data. It records the number of bytes sent to the network buffer 120 and the time taken, starting when the network buffer 120 becomes full, as detected by a blocking call to the send function. It then estimates the achievable throughput, and uses that to determine whether to switch to a higher rate

10 play stream.

As stated earlier, by knowing the data held in its network buffer 120, the server 10 implicitly knows which data has been received by the client 40 and delivered to its decoding buffer 41. This information can then be used to determine when to switch

15 between play streams, and when to invoke the error recovery procedures. However, visibility of the contents and fullness of the server's network buffer 120 in most socket implementations is not supported. In order to monitor the contents of the network buffer 120, a mirror buffer 120a is implemented. The mirror buffer 120a does not store the actual data sent to the network buffer 120, but instead stores only the number

20 of bytes sent and the timestamp of the data. Knowing the size of the network buffer 120, and assuming it is always full, the server 10 has access to the timestamp of the oldest data in the network buffer 120 via the mirror buffer 120a, which is approximately the same as the timestamp of the newest data in the client's decoding buffer 41.

25

In testing, it has been found that the assumption that the network buffer 120 at the server 10 is always full is correct at most times. This is because the transmission process is controlled to send as quickly as possible to the network buffer 120. If the network buffer 120 becomes less than full, the effect is to underestimate the amount of

30 data at the client 40, which in most cases is safe, as the major problem is seen as exhaustion of data at the client 40 rather than overflow. In practice, the decoding buffer 41 can be dimensioned to be larger than the largest amount of data it needs to

store. In any case, if the decoding buffer 41 becomes full the client 40 stops reading from the network buffer 130 which in turn stops the server network buffer 120 from emptying and transmission stops.

- 5 To determine the exact amount of data in the client's decoding buffer 41, the server also needs to know the timestamp of the data packet that the client is currently decoding and presenting. The server 10 calculates this using two assumptions: firstly that the client 40 starts decoding immediately after the server 10 sends the first packet; and secondly, that the client's clock does not drift significantly from the server's clock
10 in the duration of streaming.

- In practice both assumptions have been found to be valid. The client 40 is designed to start decoding immediately on receipt of data, and so any error on the server's estimated presentation time would result in an underestimate for the amount of data in the
15 decoding buffer 41, which as explained above is not a problem. Drift between the client's and server's clocks during a typical streaming session is most likely to be negligible compared to the amounts of data being buffered. For example, with a difference of 100 parts per million, it would take 10000 seconds, or nearly three hours, for a drift of one second to occur. In the rare case of a large amount of drift
20 accumulating, the client 40 can warn the server 10 by use of a control message, such as the one described earlier that is sent for decoding buffer underflow.

- The server 10 initially streams the play stream with the lowest bit rate, to allow the client 40 to decode and present media to the user immediately while also building up
25 the level of data in the decoding buffer 41 to provide resilience to network impairments. If the network has sufficient capacity to support transmission of a higher rate play stream, the server 10 should, at an appropriate moment in time, switch to streaming a higher rate play stream.

- 30 There are many possible strategies that could be used to determine when to switch to a higher rate play stream. Preferably, the client 40 should have sufficient data in its decoding buffer 41 to be able to continue decoding and presenting media for a

25

predetermined period of time, say 15 seconds. It is also preferred that network throughput that has been achieved in the recent past, measured over, say, the most recent 60 seconds, should be sufficient to sustain streaming of the play stream to be switched to indefinitely; that is, the recently achieved network throughput rate should
5 be greater than or equal to the bit rate of the play stream. The aim is to avoid frequent switching between streams as this can be more annoying to the user than constant quality at the lower rate.

In order to achieve this aim, it is preferred that the switching down decision includes
10 hysteresis relative to the switching up decision. For example, switching down to the next lower bit rate play stream could be triggered when the client 40 no longer has sufficient data in its decoding buffer 41 to be able to continue decoding and presenting media for a specified period of time, say 8 seconds. In the case of a configuration with three or more play streams, and the currently streamed play stream being the third or
15 even higher rate play stream, this strategy does not result in an immediate drop to the bottom of the hierarchy, as access pictures only occur periodically, and it is hoped that the decoding buffer fullness would recover after a first switch down so that a second switch down would not be necessary.

20 Figures 6a-6c are schematic diagrams of aspects of a data structure for storing an audio-visual data source suitable for use in the present invention.

The main data structure shown in Figure 6a permits the storage in a single file of multiple audio play streams, an Intra video stream, and multiple video Play and
25 Switching streams.

As the audio visual data source created and used in the present invention has a number of encoded streams that could be transmitted at any one time to a client, storage in a conventional sequential file is not possible. For example, in the case of video, a
30 particular source picture may be encoded in each play stream, and may also be encoded in the Intra stream and some or all of the Switching streams.

26

The file contains a data structure, an example of which is illustrated in Figure 6a, followed by stream data. The data structure includes a header 600 containing information about the number and type of streams (audio, video, switching etc). For the first and last instances of each type of stream it also includes pointers 610-680 (expressed as offsets from the beginning of the file) to the header for the respective stream.

Each pointer 620-680 points to a stream data structure which includes a stream header 700, containing a pointer 710 to the next stream header of the same type, a pointer 720, 730 to the first and last packets of the stream respectively. Each stream type uses a specific stream header type, however certain elements are common to all stream header types: a stream identification number 705, a pointer 710 to the next stream header of the same type and pointers 720, 730 to the first and last packets of the stream respectively. An example stream header containing only these common elements is illustrated in Figure 6b. Play and audio stream headers additionally contain the bit rate at which the stream was encoded. Switching stream headers contain the stream identifiers of the play streams from and to which the Switching stream enables switching.

Each stream consists of a sequence of packets, each represented by a packet data structure, an example of which is illustrated in Figure 6c. Each packet data structure includes a packet header 800 and a payload 810. The header includes data including a pointer 801 to the next packet in the stream, a timestamp 802, a packet sequence number 803, packet size 804, and a frame number 805 (i.e. the sequence number of the video picture or audio frame which the packet, perhaps together with other packets, represents). Switching packets additionally contain the sequence numbers of packets in from- and to- Play streams between which they allow bit rate switching to take place. The switch stream packet header effectively defines a switching point and contains the sequence number of the last packet to be played from the "from" stream before switching and the first to be played from the "to" stream after switching. Sequence numbers begin at 0, and are never negative. The use of pointers to assist in navigation

between streams when switching is possible, although this approach has not been followed in this particular embodiment.

5 The pointers to the last stream data structure and the last packet are useful when appending to a file, as they provide immediate access to the points at which the file must be extended, without the need to search through the whole file.

10 The complexity of the data structure is a consequence of packets from potentially many streams being interleaved, and of the need to support switching and recovery. Navigation from packet to packet is necessarily by pointers since, in general, packets which are consecutive within a stream will not be stored contiguously within the file. Writing of switching and recovery packets requires that precise details of source and destination packets be recorded. Switching between streams during playback requires firstly the identification of the next available switching packet, followed by playback of 15 the remaining packets from the "from" stream, playback of the switching packets, then the playback of packets from the "to" stream from the appropriate point. Furthermore there must be no appreciable delay when switching between streams.

20 In tests, both file-based and live streaming scenarios were investigated using the BTCellnet™ GPRS network. A desktop Pentium PC was used to run the encoder and Server. The client was a Compaq iPaq™ connected with via an infra-red link to a Motorola Timeport™ GPRS mobile telephone.

25 In a video-only configuration, two switching streams were used, with bit rates of 6 kbit/s and 12 kbit/s.

30 The system performed as expected. Transmission starts with the intra stream and then switches to the 6 kbit/s play stream, where it stays for some time, accumulating data in the client as a result of actually transmitting faster than 6 kbit/s. Then when sufficient data has been accumulated, and the short term average receiving rate is more than 12 kbit/s, it switches to the higher rate play stream.

At times during a lengthy session, occasional switches back to the lower rate play stream occur as a result of reduced network throughput. And very rarely, media presentation is interrupted because of a significant period during which the network could not deliver data to the client.

5

The overall effect is for most sessions, the user can view continuous media presentation, with occasional changes in quality, but no distortions of the type usually associated with bit errors and packet loss. Only very rarely are complete pauses in media presentation observed as a result of severe network impairments and loss of

10

throughput.

Claims

1. A data streaming system comprising a server (10) arranged to stream one of a plurality of encoded data streams to a client (40, 50, 60), each of the plurality of data streams being an independent representation of a common data source encoded at a different resolution to the other of the plurality of data streams, the server (10) comprising a transmitter (100) and a first buffer (120), the transmitter (100) being arranged to transmit data packets of the encoded data stream to the client (40, 50, 60) via the first buffer (120), wherein the transmitter (100) is arranged to monitor the content of the first buffer (120) and switch to transmit another of the plurality of data streams in the event that predetermined criteria are detected from the first buffer (120).
2. A system according to claim 1, wherein the transmitter (100) is arranged to determine the amount of data buffered at the client (40, 50, 60) from the content of the first buffer (120), wherein the predetermined criteria include a predetermined level of data determined to be buffered at the client.
3. A system according to claim 2, wherein a data packet is removed from the first buffer (120) upon acknowledgement by the client (40, 50, 60) of receipt of the packet.
4. A system according to claim 3, wherein the transmitter (100) is arranged to determine the amount of data buffered at the client (40, 50, 60) in dependence on the latest data packet removed from the first buffer (120) and on an estimation of number of packets decoded by the client (40, 50, 60).
5. A system according to any preceding claim, wherein the first buffer (120) includes a mirror buffer (120a) storing data on packets in the first buffer (120), the transmitter (100) being arranged to monitor the content of the first buffer (120) using the data in the mirror buffer (120a).

30

6. A system according to any preceding claim, wherein data packets are transmitted to the client (40, 50, 60) using an extended TPKT protocol, the data packets including a header containing a decoding timestamp and a data stream identifier.
- 5 7. A system according to any preceding claim, further comprising a plurality of transmitters (100), each communicating with a respective client (40, 50, 60) via a respective first buffer (120) to transmit one of the plurality of data streams determined in dependence on respective predetermined criteria.
- 10 8. A system according to any preceding claim, wherein the data stream is encoded video data.
9. A system according to claim 8, wherein the transmitter (100) is arranged to multiplex audio packets and video packets within the transmission of data packets.
- 15 10. A system according to claim 9, wherein neighbouring audio and video packets represent audio and video information that is intended for representation at substantially the same time.
- 20 11. A system according to any of claims 1 to 7, wherein the data stream is encoded audio data.
12. A system according to any preceding claim, wherein the resolution is an encoding bit rate of the data.
- 25 13. A system according to any preceding claim, wherein the server (10) includes an encoder (20) arranged to accept a data feed and encode the data feed into the plurality of encoded data streams.
- 30 14. A system according to claim 13, further comprising a plurality of buffers (70, 80, 90), wherein the encoder (20) is arranged to output each encoded data stream into a respective one of the plurality of buffers (70, 80, 90), the transmitter (100) being

31

arranged to obtain data packets for a respective data stream from its respective one of the plurality of buffers.

15. A system according to any preceding claim, wherein the server (10) includes a
5 file source (30) storing the plurality of encoded data streams.

16. A data streaming system comprising a client and a server, the server (10) being
arranged to stream one of a plurality of encoded data streams to the client (40, 50, 60),
each of the plurality of data streams being an independent representation of a common
10 data source encoded at a different resolution to the other of the plurality of data
streams, the server (10) comprising a transmitter (100) and a first buffer (120) and the
client (40, 50, 60) including a receiving buffer (130), wherein the transmitter (100) is
arranged to transmit data packets of the encoded data stream to the client (40, 50, 60)
via the first buffer (120), wherein the client (40, 50, 60) is arranged to store received
15 data packets in the receiving buffer (130) and to acknowledge receipt to the server (10),
wherein the transmitter (100) is arranged to delete packets from the first buffer (120)
when an acknowledgement receipt is received, the transmitter (100) being arranged to
switch to another of the plurality of data streams in the event that predetermined criteria
are satisfied, the predetermined criteria comprising analysis on content of the first
20 buffer (120).

17. A system according to claim 16, wherein the packets include packet sequence
data, the client (40, 50, 60) being arranged to request retransmission of non-received
packets based on the sequence data, the server (10) being arranged to retransmit a
25 packet from the first buffer (120) upon receipt of a retransmission request.

18. A method of streaming one of a plurality of encoded data streams to a client,
each of the plurality of data streams being an independent representation of a common
data source encoded at a different resolution to the other of the plurality of data
30 streams, the method comprising the steps of:
transmitting data packets of the encoded data stream to the client via a first buffer;
monitoring the content of the first buffer; and,

switching to transmit another of the plurality of data streams in the event that predetermined criteria are detected from the first buffer.

19. A method according to claim 18, wherein the plurality of data streams are each
5 encoded at a different bit rate, the method further comprising the step of initially transmitting data packets of the lowest bit rate data stream.

20. A method according to claim 18 or 19, wherein the predetermined criteria includes an amount of data determined to be buffered at client.

10

21. A method according to claim 18, 19 or 20, wherein the predetermined criteria include one or more network throughput thresholds.

22. A method according to claim 21, wherein network throughput is calculated by
15 the steps of:

counting the number of bytes passed to the first buffer;

subtracting the counted number of bytes from the size of the first buffer; and,

dividing the result by the time since the start of transmission.

20 23. A method according to claim 22, further comprising the step of measuring network throughput over more than one interval to determine throughput variation.

24. A method according to claim 22 or 23, wherein the predetermined criteria include determination of network throughput sufficient to sustain the other of the
25 plurality of the data streams.

25. A method according to any of claims 22 to 24, further comprising the step transmitting data at a maximum rate irrespective of an amount of data buffered at the client, wherein the predetermined criteria include network throughput determined at the
30 maximum rate.

33

26. A method according to any of claims 18 to 25, wherein the data streams are encoded as a series of pictures predictively encoded in dependence on the previous pictures in the data stream, the data streams including quantised source access pictures interspersed at predetermined periods in the picture series, wherein the method of
5 encoding the quantised source access pictures including the steps of:
encoding picture as a predicted picture; and,
if no information about an area of a picture is indicated in the encoded predicted picture, setting the quantiser index to a fine quantisation value when encoding as a quantised source access picture.
- 10
27. A computer program comprising computer program code means for executing the steps any of claims 17 to 24 when run on a computer.
28. A computer program as claimed in claim 27 embodied on a computer readable
15 medium.

1/6

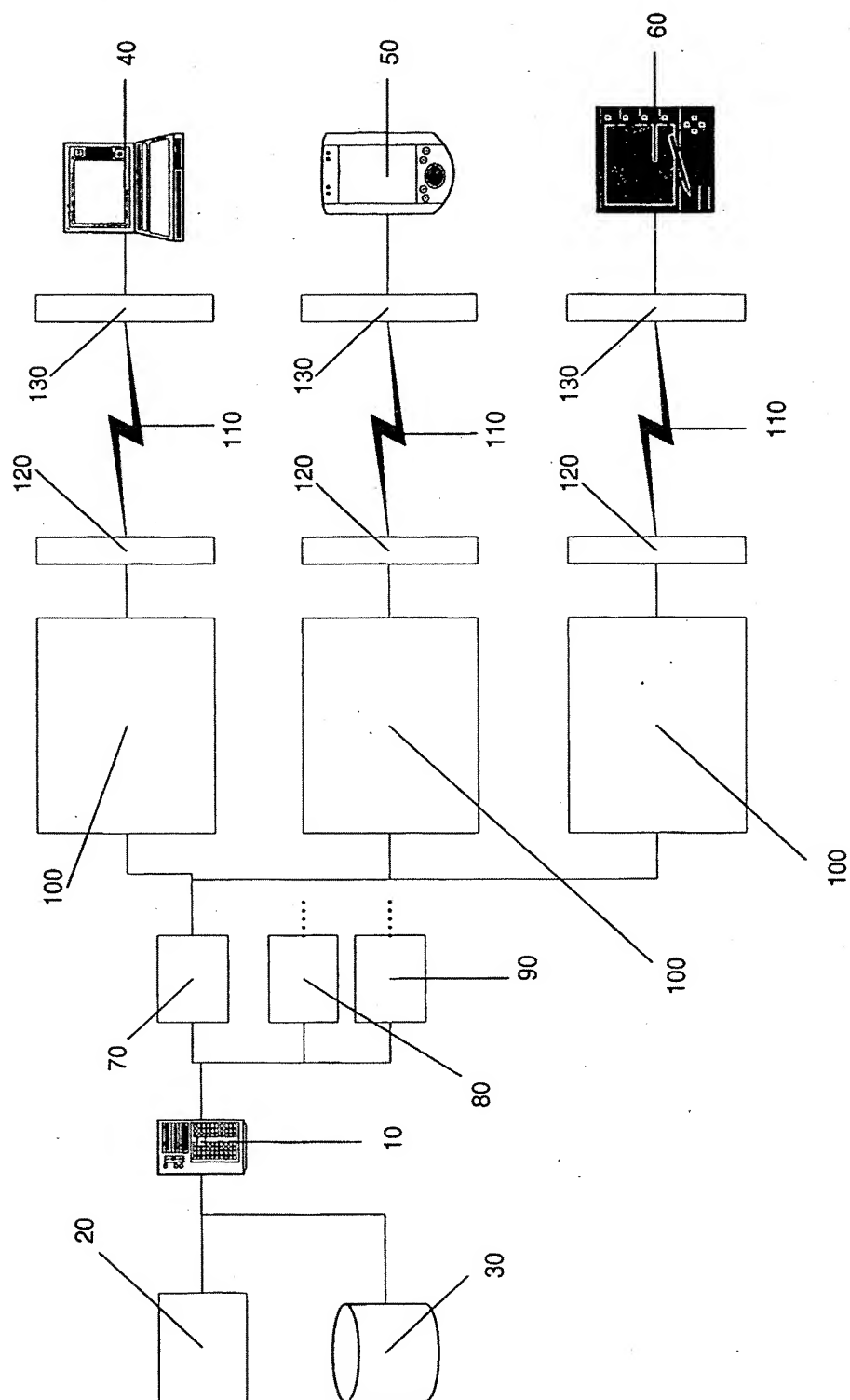


Figure 1

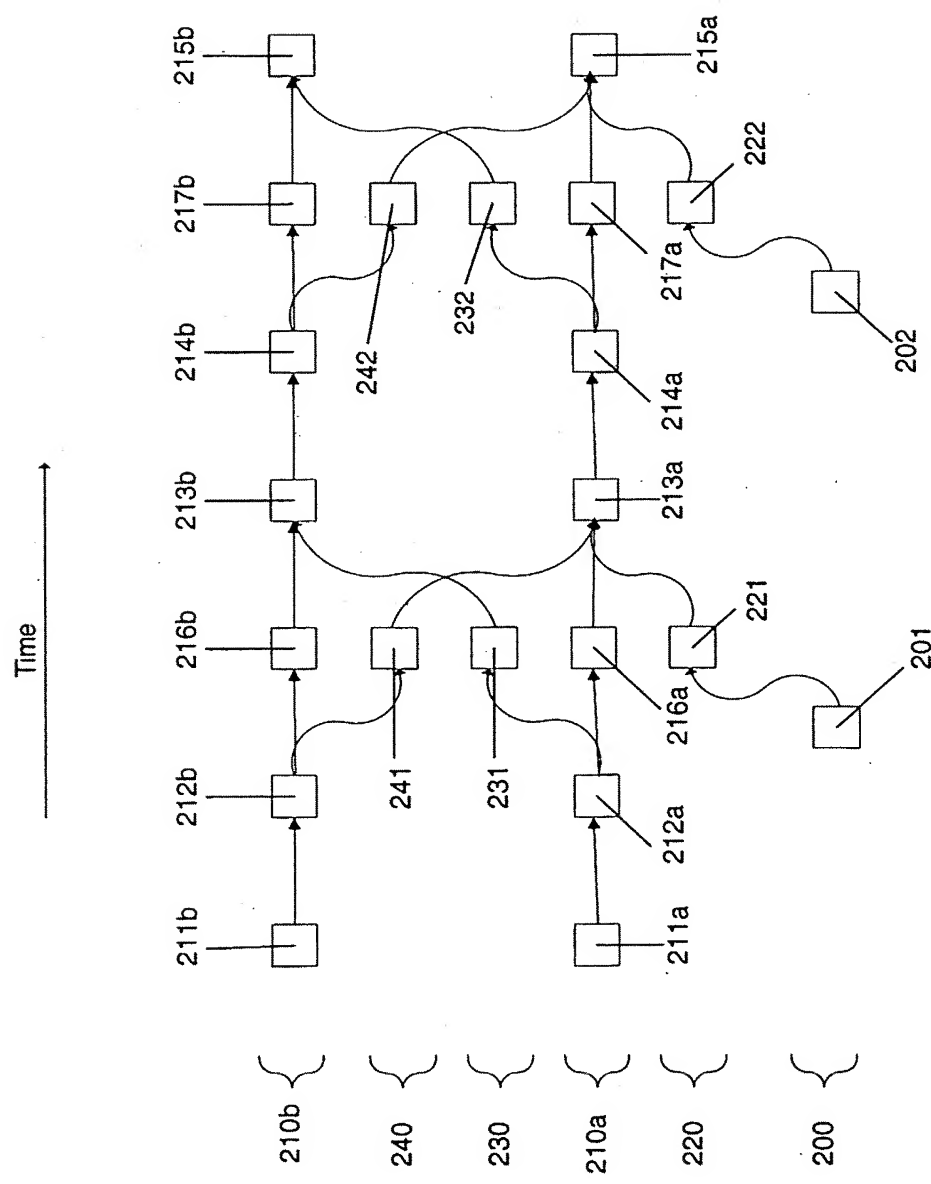


Figure 2

3/6

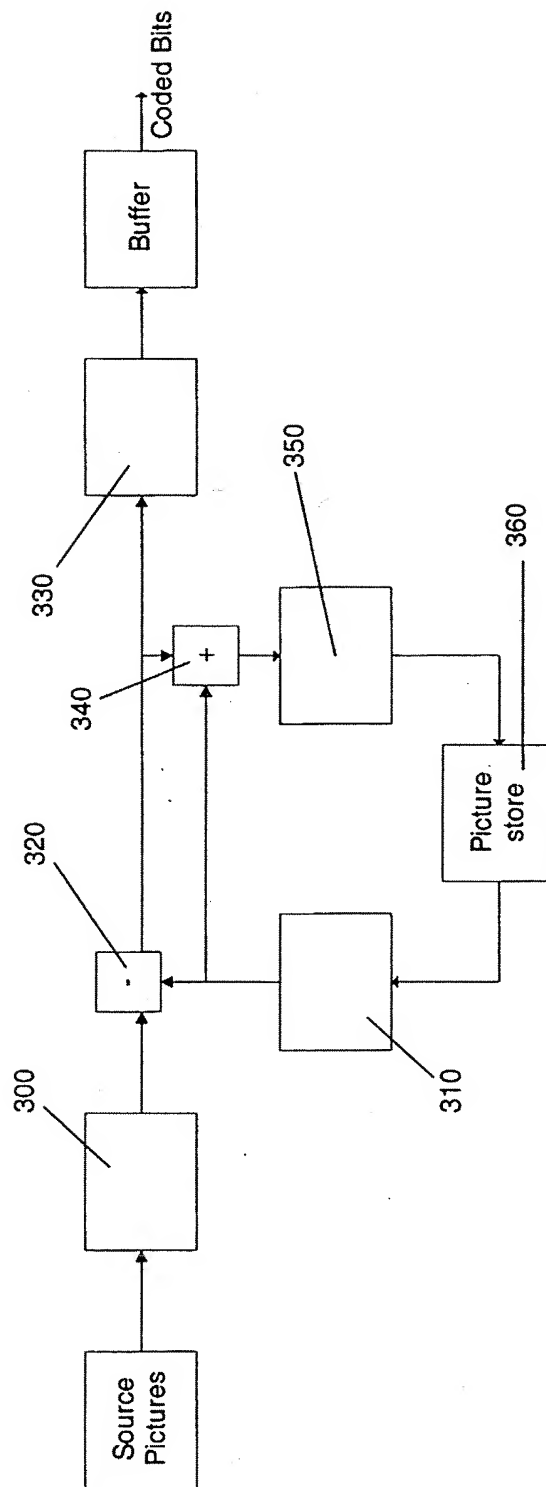


Figure 3

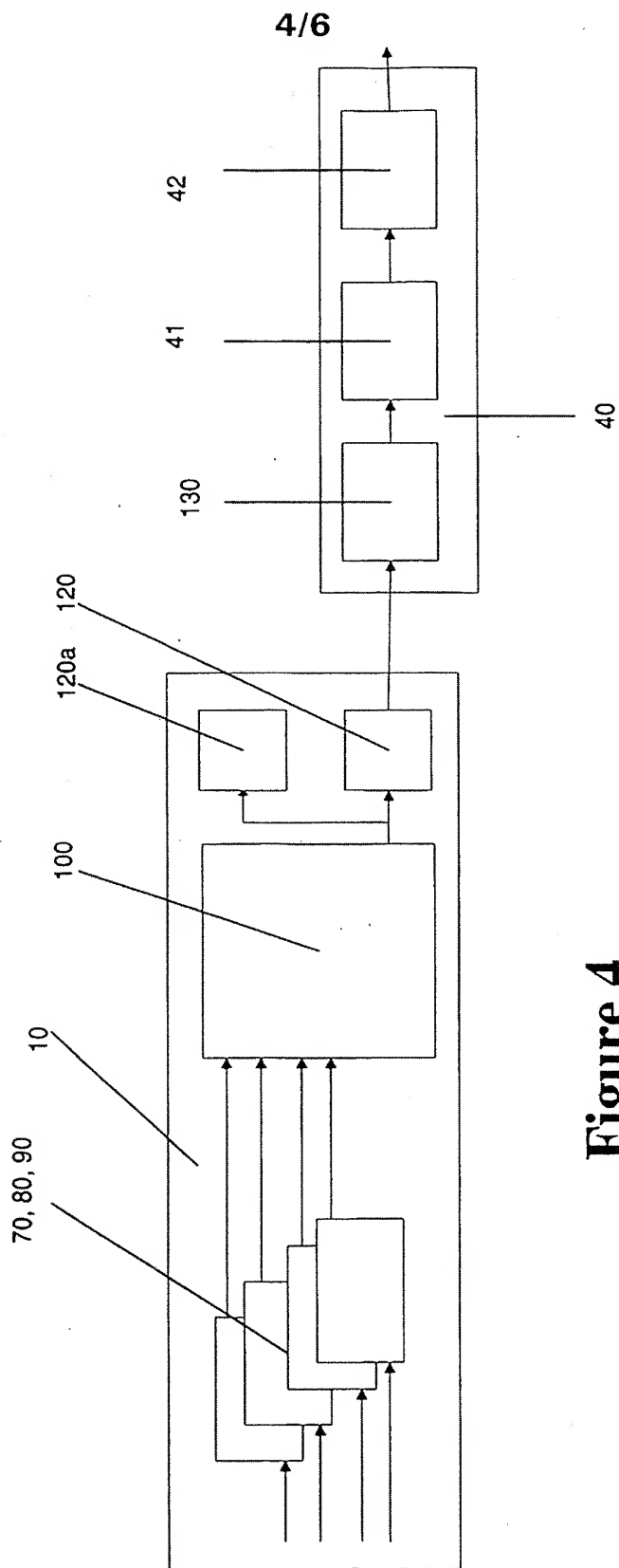


Figure 4

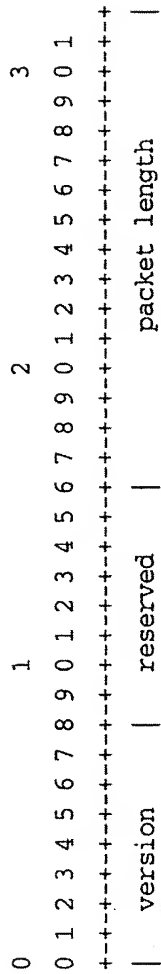


Figure 5a

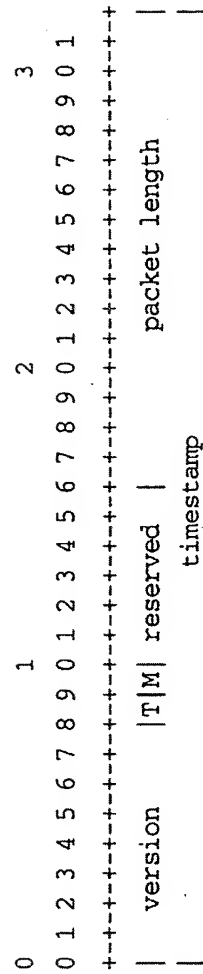


Figure 5b

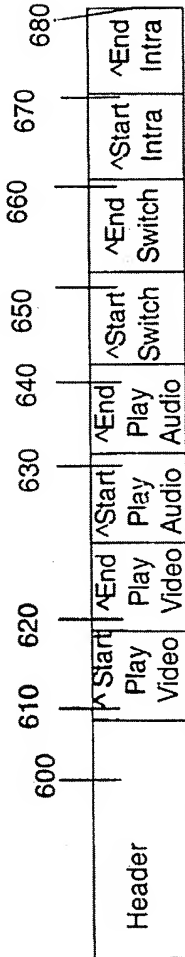


Figure 6a

007

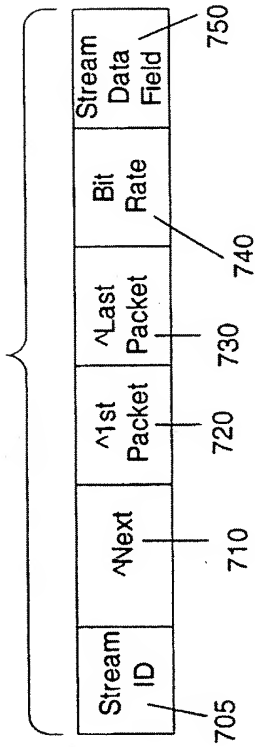


Figure 6b

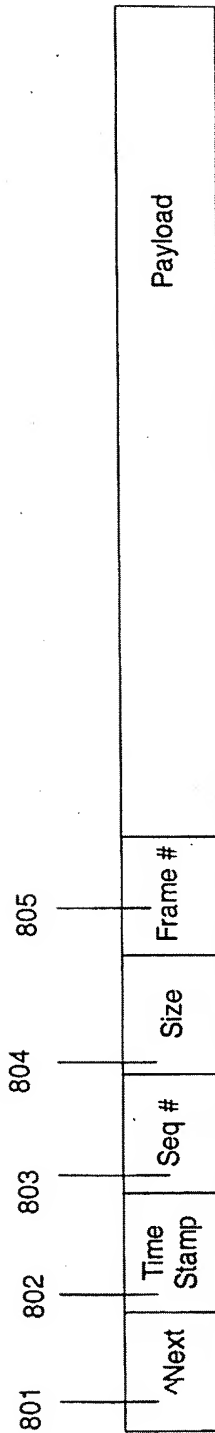


Figure 6c

008

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 03/01358

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 H04L29/06 H04N7/24

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04L H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, COMPENDEX, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 014 694 A (AHARONI AMIR ET AL) 11 January 2000 (2000-01-11)	1,5-18, 21-25, 27,28
A	column 1, line 5 - column 2, line 8 column 6, line 30 - column 8, line 41 column 10, line 22 - line 64 column 11, line 26 - column 19, line 42 figures 5-8	2-4,19, 20
X	WO 00 01151 A (SARNOFF CORP) 6 January 2000 (2000-01-06)	1,8-13, 15-19, 21-24, 27,28
A	abstract page 1, line 8 - page 3, line 15 page 8, line 11 - page 10, line 22 page 12, line 30 - page 13, line 12 ----- -/-	2,20

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"8" document member of the same patent family

Date of the actual completion of the international search

21 May 2003

Date of mailing of the international search report

15 09 2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Kesting, V.

INTERNATIONAL SEARCH REPORT

Inteq ☐ International Application No

PCT/GB 03/01358

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	REJAIE R ET AL: "Architectural considerations for playback of quality adaptive video over the Internet" IEEE INTERNATIONAL CONFERENCE ON NETWORKS. ICON. PROCEEDINGS OF ICON, XX, XX, 5 September 2000 (2000-09-05), pages 204-209, XP002177090 abstract * sections 1., 2., 3., 4. * -----	1-13, 15-25, 27,28
A	REJAIE R ET AL: "Layered quality adaptation for Internet video streaming" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, DEC. 2000, IEEE, USA, vol. 18, no. 12, pages 2530-2543, XP002214023 ISSN: 0733-8716 the whole document -----	1-5, 8-12,14, 16-25

INTERNATIONAL SEARCH REPORT

national application No.
PCT/GB 03/01358

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-25, 27, 28

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-25,27,28

System, computer readable medium and method for switching from a first of a plurality of data streams transmitted from a server via a server buffer to a client, to another of the plurality of data streams based on a determination, made by the server from the content of the buffer, of the amount of data buffered at the client

2. claim: 26 (as dependent on claim 18)

Method for streaming, from a server to a client, encoded data streams including quantized source access pictures wherein during encoding of an access picture, a quantiser index is set to a fine quantisation value when no information about an area of the picture is indicated in the result of a predictive encoding of the picture

INTERNATIONAL SEARCH REPORT

In
nal Application No
PCT/GB 03/01358

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 6014694	A	11-01-2000	AU 7928998 A	19-01-1999
			CA 2294511 C	02-04-2002
			EP 0992160 A1	12-04-2000
			WO 9900984 A1	07-01-1999
			JP 2002511216 T	09-04-2002
WO 0001151	A	06-01-2000	WO 0001151 A1	06-01-2000